08/746,581

대 한 민 국 특 허 청
KOREAN INDUSTRIAL PROPERTY OFFICE

별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Industrial Property Office.

출 원 번 호 : 1995 년 특허출원 제 53941 호
Application Number

출 원 년 월 일 : 1995 년 12 월 22 일
Date of Application

출 원 인 : 한국전자 통신 연구소
Applicant(s)

1996 년 3 월 5 일

특 허 청
COMMISSIONER

정 본

접수 인란		방식 심사란	

특허청 출원과 출납 1996. 2. 12 제 접수 회

심 96 2 12 상통

최 96 2 13 덕철

009924

보 정 서

제출인 (출원인)	성 명	재단법인한국전자통신연구소 소 장 양 승 택				
	사건과의관계	출 원 인		국 적	대한민국	
	주 소	대전직할시 유성구 가정동 161				
대리인	성 명	박 해 천	대리인코드	F196	전화번호	555-7503
	주 소	서울시 강남구 역삼1동 740-5 동방빌딩 1층				
사건의표시	출원번호	1995년 특허출원 제 53941 호			출원일자	1995. 12. 22.
발명의 명칭		다층구조 신경망을 이용한 음소 분할 방법				
제출원인		☑ 자 진 ☐ 통 지 (또는 명령) : 통지를 받은 날짜 (. .) 제출마감날짜 (. .)				
보정할사항		☐ 출원서(발명의 명칭) ☑ 상세한 설명 ☑ 특허청구의 범위 ☐ 위임장 ☐ 도면 ☐ 우선권주장일자, 번호 ☐ 대표자명 ☑ 추가심사청구료 ☐ 수수료 ☐ 청구이유 ☐ 기타				
보정내용 및 이유		(별지 사용)				
심사청구상황		☐ 심사미청구 ☑ 심사청구 : 청구일자 (1995. 12. 22.)				
수 수 료		☐ 보정료 ( )원		☑ 추가심사청구료 (34,000)원		

보정에 의한 청구항수 및 청구료의 증가

산출 대상 구분	최초출원시 청구범위항수	1차 보정시 청구 범위 항수			2차 보정시 청구 범위 항수		
최종항번호	1	3					
독립항수	1	삭제	신설	청구항수	삭제	신설	청구항수
				1			
종속항수	0		2	2			

특허법시행규칙 제 13 조의 규정에 의하여 위와 같이 제출합니다.

1996년 2 월 12 일

대리인 변리사 박 해 천

특허청장 귀하

구비서류

1. 보정서 부본 2통.
2. 명세서 정본 1통, 부본 2통.

명 세 서

1. 발명의 명칭

다층구조 신경망을 이용한 음소 분할 방법

2. 도면의 간단한 설명

제 1 도는 본 발명이 적용되는 시스템의 구성도,

제 2 도는 본 발명에 이용되는 다층 신경망의 구성도,

제 3 도는 본 발명의 일실시예에 따른 전체 흐름도,

* 도면의 주요부분에 대한 부호의 설명

1 : 음성 입력부

2 : 전처리부

3 : MLP 음소 분할부

4 : 음소 경계 출력부

3. 발명의 상세한 설명

본 발명은 다층구조 신경망을 이용한 음소 분할 방법에 관한 것이다.

종래의 음소 분할 기술은 음성 신호로부터 주파수 성분인 스펙트로그램(spectrogram)을 추출한 후 사전에 정해진 여러가지 음성학적 지식과 규칙을 적용한 분석을 통해 음소의 경계를 찾아내는 방법을 사용함으로써 시스템의 복잡도가 매우 높은 문제점이 있었다.

또한, 음소 분할을 위해 사용하는 여러가지 지식과 규칙 상호간의 효율적이고도 최적의 결합 방법이 없기 때문에 실제 사용시에는 시스템의 성능이 신뢰할 정도가 되지 못한다는 점과 실제 사용될 때의 상황 변화에 따라 시스템의 성능이 급격히 저하된다는 문제점이 있었다.

다른 방법으로는 음소 분할을 위해서 모든 음소들의 특징을 사전에 추출하여 패턴으로 저장한 후, 음소 분할시에 모든 음소들에 대한 특징

패턴을 번갈아가면서 입력된 음성 신호와 비교하여 음소의 경계를 찾아 내는 기술을 들수 있다.

이 방법은 모든 음소들에 대한 특징 패턴의 정보를 가지고 있어야 하므로 시스템의 메모리 양이 커지게 되고 수행과정에서의 계산량도 증가함으로써 경제적인 시스템을 구현하지 못한다는 문제점이 있었다.

따라서, 상기의 문제점을 해결하기 위하여 안출된 본 발명은 음소 자체에 대한 부가적인 지식없이 음소와 음소의 경계에서 나타나는 음성 신호상의 변화만을 이용하여 음소의 경계가 되는 지점을 정확하고도 효율적으로 포착하여 자동 음소 분할이나 음소 레이블링이 필요한 응용 분야에 유익하게 활용될 수 있는 다층구조 신경망을 이용한 음소 분할 방법을 제공하는 데 그 목적이 있다.

상기 목적을 달성하기 위한 본 발명은 발성된 음성으로 부터 디지탈로 변환된 음성 샘플을 출력하는 음성 입력부, 상기 음성 입력부로 부터 입력된 음성 샘플로 부터 음소 분할에 적합한 특징 벡터를 추출하는 전처리부, 상기 전처리부의 특징 벡터를 이용하여 음소의 경계 부분을 찾아 출력하는 다층 신경망(MLP : Multi Layer Perceptron) 음소 분할부,

및 상기 MLP 음소 분할부의 음소의 경계에 관한 위치 정보를 프레임 위치의 형태로 출력하는 음소 경계 출력부를 구비한 음소 분할 장치에 적용되는 다층구조 신경망을 이용한 음소 분할 방법에 있어서, 디지탈화된 음성 샘플들로부터 음성을 연속적으로 세그먼트화하여 음성을 프레임화하고, 각 음성 프레임들에 대하여 프레임별 특징 벡터를 추출한 후, 프레임별 특징 벡터들의 인접 프레임간 차이를 구한 프레임간 특징 벡터를 추출하여 특징 벡터들의 최대치와 최소치를 정규화하는 제 1 단계; 상기 제 1 단계 수행 후, 다층 신경망(MLP)의 입력층과 은닉층, 은닉층과 출력층 사이에 존재하는 가중 함수를 초기화하고, 다층 신경망(MLP)의 출력 목표 데이타를 지정하고, 특징 벡터를 다층 신경망(MLP)에 입력하여 학습시킨 후, 평균자승오차의 감소 비율이 허용 한계내로 수렴되면 학습을 통해 구한 가중 함수와 MLP의 규격에 대한 정보를 저장하고 종료하는 제 2 단계; 및 상기 제 2 단계에서 구한 가중 함수를 읽고, 특징 벡터를 입력받아 음소 경계 판정을 위한 연산을 수행하여 출력값을 생성한 후, 출력값에 따라 음소 경계 부분을 판정하고, 현재의 분석 프레임이 입력된 음성의 최종 프레임의 임의의 프레임 이전에 도달하였으면, 구해진

음소의 경계를 프레임 번호로 나타낸 값을 최종 결과로 출력하는 제 3 단계를 포함하는 것을 특징으로 한다.

이하, 첨부된 도면을 참조하여 본 발명의 일실시예를 상세히 설명한다.

제 1 도는 본 발명이 적용되는 시스템의 구성도로서, 도면에서 1은 음성입력부, 2는 전처리부, 3은 다층 신경망(MLP : Multi Layer Perceptron) 음소 분할부, 4는 음소 경계 출력부를 각각 나타낸다.

음성입력부(1)는 공기중의 음성 파형으로 부터 전기적인 음성 신호로 변환하는 마이크와 전기적인 아날로그 신호로 입력된 음성 신호에서 저주파 잡음과 고주파 에일리어싱(aliasing) 효과를 제거하기 위한 대역 통과 여파기(band pass filter), 및 아날로그 음성 신호를 디지탈 음성 신호로 변환하는 아날로그/디지탈 변환기(ADC : Analog to Digital Converter)로 구성되며, 그 기능은 발성된 음성으로 부터 디지탈로 변환된 음성 샘플을 얻어 전처리부(2)에 출력한다.

상기 전처리부(2)는 상기 음성 입력부(1)에서 입력된 음성샘플들로 부터 음소 분할에 적합한 특징 벡터를 추출하여 MLP 음소 분할부(3)로 출

력한다.

상기 MLP 음소 분할부(3)는 상기 전처리부(2)로 부터 입력된 특징 벡터들을 사용하여 음소의 경계 부분을 찾아 음소 경계 출력부(4)로 출력한다.

상기 음소 경계 출력부(4)는 상기 MLP 음소 분할부(3)에서 자동적으로 분할된 음소의 경계에 관한 위치 정보를 프레임 위치의 형태로 출력한다.

제 2 도는 본 발명에 이용되는 다층 신경망의 구성도를 나타낸다.

본 발명은 종래의 지식이나 규칙 기반의 음소분할 방법의 단점을 보완하기 위하여 효과적이고도 신뢰성있는 자동 음소 분할기(phoneme segmenter)를 신경망의 한 종류인 다층 신경망(MLP)을 사용하여 구현하였다.

MLP를 이용하는 음소 분할 방법은 종래의 음소분할 방법들의 문제점으로 알려진 음성 신호에 내재되어 있는 음소 경계에 관한 지식이나 규칙의 불완전한 모델링에서 오는 성능 저하를 해결하는데 매우 적합하다. 즉, 많은 음성 데이타에서 추출한 특징 벡터로 부터 음소 분할에 필요한

기능을 학습을 통해서 스스로 배우도록 함으로써 음소의 경계에 대한 특별한 가정이나 규칙 및 지식을 사전에 도입하지 않고도 음성신호 자체에 내재된 지식이나 규칙을 MLP로 하여금 스스로 찾아내도록 하는 방법이다. 그러므로, 본 발명은 음성 신호의 모델링을 용이하게 하기 위하여 사전에 그 분포나 모델링을 위한 불확실한 가정의 도입이나 추가적인 처리를 할 필요가 없는 장점이 있다.

본 발명에 이용되는 다층 신경망(MLP)의 구조는 입력(input), 은닉(hidden), 출력(output)의 세 가지 층(layer)으로 구성된 다층 구조의 형태를 취하고 있다.

도면에서와 같이 하단에 위치한 입력층은 연속적인 다섯 프레임에서 발생하는 4개의 인접 프레임 간 차이로 부터 추출된 총 72개의 프레임간 특징 벡터들에 대한 입력 노드들과 다층 신경망(MLP)의 은닉층에서의 문턱치 비교 과정 대신에 사용되는 입력값 1을 위한 입력 노드 한개를 포함하여 모두 73개의 입력 노드로 구성되어 있다.

출력층의 출력 노드는 음소의 경계임을 나타내는 첫번째 노드와 그렇지 않은 경우를 나타내는 두번째 노드를 합하여 모두 2개로 구성되어 있

으며, 입력층과 출력층의 사이에 위치한 은닉층은 다층 신경망(MLP)이 실제로 구현해야 하는 비선형 분리(nonlinear discrimination)기능이 이루어지는 계층이다.

이 은닉층의 활성화 함수(activation function)로 다음과 같은 비선형의 S자 모양(sigmoid)의 함수를 사용한다.

$$y = (\exp(x) - 1) / (\exp(x) + 1),$$

여기서, x, y는 각각 활성화 함수의 입력과 출력을 나타낸다.

은닉층의 노드 수 N은 다층 신경망(MLP)의 최종 성능과 밀접한 관련이 있다고 알려져 있는데 여러가지 데이타를 사용한 실험을 통해서 10에서 30 사이가 적당하다.

입력층과 은닉층, 은닉층과 출력층 사이에는 각 층의 노드들을 전부 연결하는 가중함수(weight)들이 존재한다. 이 가중 함수들은 층과 층 사이의 노드들을 전부 연결시키기 때문에, 입력층과 은닉층의 경우에는 그 수가 입력노드의 수 x 은닉노드의 수 = 73 x N 개가 있으며, 은닉층과

출력층의 경우에는 은닉노드의 수 x 출력노드의 수 = N x 2 개가 존재한다. 이 가중 함수들은 오류 역전파 알고리즘을 이용한 학습을 통해서 사전에 구해진 다음 메모리에 저장되어 있다가 음소 분할시에 불러내어 사용된다.

제 3 도는 본 발명의 일실시예에 따른 전체 흐름도로서, 전처리부(2)와 MLP 음소 분할부(3)의 내부에서 음소 분할 알고리즘의 동작 과정을 나타낸 것으로 MLP 음소 분할 알고리즘의 학습 과정과 분할 과정의 2 부분으로 구성되어 있다.

먼저, 음성 프레임화와 특징 벡터 추출 과정은 전처리부(2)에서 수행되는 과정으로서 학습과 분할 두 과정에 공통적으로 사용된다.

본 발명에서의 특징 벡터들의 선정에서는 음소간의 경계에서 음성 스펙트럼의 변화가 심하다는 점을 이용하기 위하여 각 프레임간의 스펙트럼의 차이를 잘 나타내주는 인자를 유도하였다.

먼저, 디지탈화된 음성 샘플들로부터 음성의 특징을 추출하기에 알맞은 길이로 음성을 연속적으로 세그먼트화하여 음성을 프레임화 한다 (10).

음성 프레임화는 입력된 전체 음성 샘플들에 대해서 매 10 msec마다 16 msec의 길이로 해밍(Hamming) 창함수(window)를 취하여 음성 프레임화한다.

다음은 음성 프레임으로 부터 특징 벡터를 추출하는데 첫단계에서는 앞에서 구해진 각 음성 프레임들에 대하여 음성의 특징을 효과적으로 잘 나타내는 프레임별 특징 벡터들을 음성학적인 지식에 근거하여 추출한다.

그런다음, 두번째 단계에서는 첫단계에서 구한 프레임별 특징 벡터들에 대하여 인접 프레임간 차이를 구한 프레임간 특징 벡터를 추출하여 이들을 MLP 음소 분할부(3)에 입력되는 최종적인 특징 벡터로 사용한다(11).

상기 과정을 보다 상세히 설명하면, 먼저 1차적으로 각각의 프레임들에 대해서 구한 특징 벡터는 다음과 같다.

(1) 프레임 에너지 : 음성의 프레임별 발성강도를 나타내는 것으로서 다음과 같이 구하였다.

$$ENG\_FRM(t) = \log 10(\sum_{n} s(n)*s(n)) \ , \ n=0,1, \ ...,N,$$

여기서 s(n)은 t번째 프레임에 속한 음성 샘플을 나타내고, N은 음성 프레임의 길이를 나타낸다.

(2) 16차 멜 스케일 FFT(mel-scaled fast Fourier transform) : 프레임별 음성의 주파수 특성인 스펙트럼을 구하기 위하여 먼저 FFT(fast Fourier transform)를 한 후 얻어진 음성의 주파수 성분을 인간의 청각 특성과 유사하게 사전에 정해진 16 개의 주파수 대역으로 분류한 16차의 대역별 에너지를 구하여 멜 스케일 FFT 계수로 사용한다. 프레임 인덱스 t에서 j차 멜 스케일 FFT 계수는 다음 식과 같이 구해진다.

$$MSFC(j, t) = \log 10(\sum_{f=1}^{16} s(j,t,f)),$$

f는 각 주파수 밴드에 포함된 주파수, 여기서 j는 각 주파수 대역의 인덱스를 나타내고 s(j,t,f)는 FFT로부터 구해진 t번째 프레임의 j차 주파수 대역 진폭 스펙트럼의 주파수별 성분을 나타낸다.

(3) 대역별 에너지 비 : 음소 분할시에 유성음과 무성음으로 된 음

소를 정확하게 구분하는 일이 매우 중요한데 이 유,무성음의 큰 차이점은 에너지의 주파수 대역별 분포이다. 따라서, 본 발명에서는 유,무성음의 구분을 위해 0-3kHz 사이에 존재하는 저주파 에너지와 3 kHz - 8 kHz 사이에 분포하는 고주파 에너지를 각각 구한다음 이들의 비를 특징 벡터의 하나로 선정하였다.

$$ENG\_RTO(t) = \log10(ENG\_LOW(t)) - \log10(ENG\_HIGH(t))$$

$$ENG\_LOW(t) = \sum_f s(f, t), \quad f=0,..., 3\ kHz.$$

$$ENG\_HIGH(t) = \sum_f s(f, t), \quad f=3kHz, ..., 8\ kHz.$$

여기서 ENG_LOW(t), ENG_HIGH(t)는 각각 t 번째 음성 프레임의 저주파대와 고주파대의 에너지로서 FFT에서 구한 진폭 스펙트럼에서 각 대역에 포함된 성분들의 합으로 구한다.

최종적인 MLP 음소 분할부(3)의 입력으로 사용되는 프레임간 특징 벡터는 음소 분할이 음소간의 경계에서 큰 변화를 나타낸다는 특징에 근거하여 위에서 구한 일차적인 프레임별 특징 벡터들에 대해서 인접 프레

임간의 차이를 다음과 같이 구함으로써 얻는다.

(1) 프레임 에너지의 인접 프레임간 차이

$$dENG\_FRM(t) = |ENG\_FRM(t) - ENG\_FRM(t-1)|$$

(2) 16차 멜 스케일 FFT의 프레임간 차이

$$dMSFC(j,t) = |MSFC(j,t) - MSFC(j, t-1)|, \quad j=0,1, ..., 15.$$

여기서 j는 계수들의 각 차수를 나타낸다.

(3) 대역별 에너지 비의 프레임간 차이

$$dENG\_RTO(t) = |ENG\_RTO(t) - ENG\_RTO(t-1)|$$

이렇게 특징 벡터를 추출한 후, MLP 음소 분할부(3)의 입력으로 사

용하기 위해 특징 벡터들의 최대치와 최소치가 각각 1과 -1이 되도록 정규화(normalize) 한다(12).

이렇게 정규화된 특징 벡터를 이용한 MLP 음소 분할부(3)의 학습 과정을 살펴보면, MLP 음소 분할부(3)의 학습하기 위한 초기 단계로서 입력층과 은닉층, 은닉층과 출력층 사이에 존재하는 가중 함수(weight)를 초기화한다(13). 초기치는 1과 -1 사이에 분포하는 무작위의 값으로 선정하였다.

그런 다음 음소의 경계 부분을 찾도록 가르치는 출력층의 출력 목표 데이타를 지정한다(14). 각 프레임별로 출력 목표 데이타는 MLP 출력 노드의 수와 같은데, 음소의 경계일 경우(1,-1) 경계가 아닐 경우 (-1,1)의 값을 갖는다. 이 출력 목표 데이타는 사전에 음소 분할된 음성 데이타베이스로 부터 구한 음소의 경계 정보를 이용하여 해당 특징 벡터의 프레임 위치와 일치되도록 작성된다.

이렇게 출력 목표 데이타를 지정한 후, 학습 데이타인 특징 벡터를 MLP의 입력층에 입력하여(15), MLP를 학습 시킨다(16). 입력층에는 연속하는 4개의 프레임간 특징벡터의 입력을 위한 72개의 입력 노드와 은

닉층의 문턱값 비교 과정 대신에 입력되는 1을 위한 하나의 입력 노드를 합하여 전체 73개의 노드로 구성된다.

4개의 프레임간 특징 벡터들은 제 2 도에 나타낸 하단에서와 같이 현재 분석 프레임 t를 중심으로 전후 2 프레임(t-2, t-1, t+1, t+2)씩을 포함한 5 프레임으로 부터 발생하는 4개의 프레임 사이에서 각각 추출된다.

음소분할 MLP의 학습 알고리즘은 일반적으로 사용하는 오류역전파(error back propagation) 알고리즘을 사용한다.

이렇게 MLP를 학습시킨 후, 평균자승오차(mean squared error)의 감소비율이 허용한계 내로 수렴되었으면(17) 학습을 통해서 구해진 가중함수들과 MLP의 규격에 대한 정보를 저장한 후(18) 학습 과정을 종료한다.

학습 과정을 종료한 후 상기에서 설명한 바와 같이 디지탈화된 음성 샘플들로부터 음성의 특징을 추출하기에 알맞은 길이로 음성을 연속적으로 세그먼트화하여 음성을 프레임화 한 후(10), 특징 벡터를 추출하고(11), 이를 정규화 한다(12).

그런 다음 상기 학습 과정에서 구해진 가중 함수들을 MLP 은닉층으

로 읽어들이고(19), 상기 과정에서 구한 특징벡터 72개를 MLP의 입력노드의 순서에 따라서 입력시키고, 마지막 73번째 입력노드에 1을 입력한다(20).

MLP 음소 분할부(3)에서는 입력된 특징 벡터들에 대하여 다음과 같은 MLP 연산을 통해서 음소 경계 판정을 위한 출력값을 생성한다(21).

$$HID(j) = SGMOD(\sum_{i} IN(i) \times WGT\_IH(i,j)),\ i=0,1,...,72.\ j=0,1,...,N-2,$$

$$HID(N-1) = 1,$$

$$OUT(k) = SGMOD(\sum_{j} HID(j) \times WGT\_HO(j,k)),\ j=0,1,...,N-1,\ k=0,1.$$

여기서 IN(j)는 i 번째 입력 노드의 입력을, OUT(k)는 k번째 출력 노드의 출력을, WGT_IH(i,j)는 i번째 입력 노드와 j번째 은닉 노드를 연결하는 가중 함수를, WGT_HO(j,k)는 j번째 은닉 노드와 k번째 출력 노드를 연결하는 가중 함수를 나타내며, SGMOD는 전술한 S자 모양(sigmoid) 함수를 나타낸다. 또한 최종 출력 노드에서의 문턱값 비교 과정을 대신하기 위해 마지막 은닉 노드에 1을 지정한다.

다음 음소 경계 부분을 판정하는 데 앞의 MLP 음소 분할부(3)에서 연산된 출력값을 비교하여 첫번째 출력값인 OUT(0)이 양수이면 그 분석 프레임이 음소의 경계이고, 반대로 OUT(1)이 양수이면 음소의 경계가 아닌 것으로 판정한다(22).

그런 후, 현재의 분석 프레임이 입력된 음성의 최종 프레임의 2 프레임 이전에 도달하였나를 검사하여(23) 도달하지 않았으면 MLP 입력층에 특징 벡터를 입력하는 이하의 과정을 반복하고, 도달하였으면 구해진 음소의 경계를 프레임 번호로 나타낸 값을 최종 결과로 출력하고(24), 종료한다.

상기와 같이 동작하는 본 발명은 인간과 기계사이의 대화를 가능하게 해주는 기술인 음성 인식 시스템의 구현에 있어서 먼저 음소 단위로 음성을 분할한 다음 분할된 음소 세그먼트에 대하여 음소인식을 수행하는 음소분할 기반의 음소 인식에 필수적인 정확하고 효율적인 음소분할 전처리를 가능하게 해주며, 음소 단위의 음성 인식 및 음성합성 시스템의 구현에 필요한 다량의 음소 분할된 음성 데이타베이스를 구축할 때도 지금까지의 음성전문가에 의한 수작업을 대신하여 신뢰성과 일관성있게 자

동적인 음성분할을 가능하게 함으로서 많은 시간과 비용의 절감을 가져 오는 효과가 있다.

4. 특허 청구의 범위

1. (정정) 발성된 음성으로 부터 디지탈로 변환된 음성 샘플을 출력하는 음성 입력부(1), 상기 음성 입력부(1)로 부터 입력된 음성 샘플로 부터 음소 분할에 적합한 특징 벡터를 추출하는 전처리부(2), 상기 전처리부(2)의 특징 벡터를 이용하여 음소의 경계 부분을 찾아 출력하는 다층 신경망(MLP : Multi Layer Perceptron) 음소 분할부(3), 및 상기 MLP 음소 분할부(3)의 음소의 경계에 관한 위치 정보를 프레임 위치의 형태로 출력하는 음소 경계 출력부(4)를 구비한 음소 분할 장치에 적용되는 다층구조 신경망을 이용한 음소 분할 방법에 있어서,

디지탈화된 음성 샘플들로부터 음성을 연속적으로 세그먼트화하여 음성을 프레임화하고, 각 음성 프레임들에 대하여 프레임별 특징 벡터를 추출한 후, 프레임별 특징 벡터들의 인접 프레임간 차이를 구한 프레임간 특징 벡터를 추출하여 특징 벡터들의 최대치와 최소치를 정규화하는 제 1 단계(10 내지 12);

상기 제 1 단계(10 내지 12) 수행 후, 다층 신경망(MLP)의 입력층과

은닉층, 은닉층과 출력층 사이에 존재하는 가중 함수를 초기화하고, 다층 신경망(MLP)의 출력 목표 데이타를 지정하고, 특징 벡터를 다층 신경망(MLP)에 입력하여 학습시킨 후, 평균자승오차의 감소 비율이 허용 한계내로 수렴되면 학습을 통해 구한 가중 함수와 MLP의 규격에 대한 정보를 저장하고 종료하는 제 2 단계(13 내지 18); 및

상기 제 1 단계(10 내지 12) 및 상기 제 2 단계(13 내지 18) 수행 후, 상기 제 2 단계(13 내지 18)에서 구한 가중 함수를 읽고, 특징 벡터를 입력받아 음소 경계 판정을 위한 연산을 수행하여 출력값을 생성한 후, 출력값에 따라 음소 경계 부분을 판정하고, 현재의 분석 프레임이 입력된 음성의 최종 프레임의 2 프레임 이전에 도달하였으면, 구해진 음소의 경계를 프레임 번호로 나타낸 값을 최종 결과로 출력하는 제 3 단계(19 내지 24)를 포함하는 것을 특징으로 하는 다층구조 신경망을 이용한 음소 분할 방법.

2. (신설) 제 1 항에 있어서,

상기 제 1 단계(10 내지 12)의 음성 프레임화 과정은, 입력된 전체 음

성 샘플들에 대해서 10 msec마다 16msec의 길이로 해밍 창함수를 취하여 수행되는 것을 특징으로 하는 다층구조 신경망을 이용한 음소 분할 방법.

3. (신설) 제 1 항에 있어서,

상기 제 3 단계(19 내지 24)의 음소 경계 부분의 판정은 연산을 수행하여 생성된 출력값을 비교하여 첫번째 출력값인 OUT(0)이 양수이면 그 분석 프레임이 음소의 경계이고, 반대로 OUT(1)이 양수이면 음소의 경계가 아닌 것으로 판정하는 것을 특징으로 하는 다층구조 신경망을 이용한 음소 분할 방법.

요 약 서

본 발명은 다층구조 신경망을 이용한 음소 분할 방법에 관한 것으로서, 음성 입력부(1), 전처리부(2) 다층 신경망(MLP : Multi Layer Perceptron) 음소 분할부(3), 및 음소 경계 출력부(4)를 구비한 음소 분할 장치에 적용되는 다층구조 신경망을 이용한 음소 분할 방법에 있어서, 디지탈화된 음성 샘플들로부터 음성을 프레임화하고, 각 음성 프레임들에 대하여 프레임별 특징 벡터를 추출한 후, 프레임별 특징 벡터들의 인접 프레임간 차이를 구한 프레임간 특징 벡터를 추출하여 특징 벡터들의 최대치와 최소치를 정규화하고, 학습을 통해 가중 함수와 MLP의 규격에 대한 정보를 구해 저장하고, 상기 과정에서 구한 가중 함수를 읽고, 특징 벡터를 입력받아 음소 경계 판정을 위한 연산을 수행하여 출력값에 따라 음소 경계 부분을 판정하고, 현재의 분석 프레임이 입력된 음성의 최종 프레임의 2 프레임 이전에 도달하였으면, 구해진 음소의 경계를 프레임 번호로 나타낸 값을 최종 결과로 출력하는 과정을 통해 음소 자체에 대한 부가적인 지식없이 음소와 음소의 경계에서 나타나는 음성 신호상의

변화 만을 이용하여 음소의 경계가 되는 지점을 정확하고도 효율적으로 포착하여 자동 음소 분할이나 음소 레이블링이 필요한 응용 분야에 유익하게 활용될 수 있는 효과가 있다.

정 본

IPC 분류기호	분류		방식심사란	출원번호 :		
	부분류			담 당	심사관	53941

특 허 출 원 서 (b)

접수인	출원인	성 명	한국전자통신연구소 소 장 양 승 택		출원인코드	37500300
		주 소	대전시 유성구 가정동 161번지		우편 번호	305-350
대리인	성 명	박 해 천	변리사 등록 번호	F196	전화번호	555-7503
		염 주 석		H386		
	주 소	서울특별시 강남구 역삼1동 740-5 동방빌딩 1층				
발명자	성 명	이 영 직	주민등록번호	560824-1047217	국 적	대한민국
	주 소	대전시 유성구 어은동 99 한빛아파트 102동 105호				
	성 명	서 영 주	주민등록번호	691216-1788216	국 적	대한민국
	주 소	대전시 유성구 가정동 236-1				
	성 명	양 재 우	주민등록번호	520709-1177710	국 적	대한민국
	주 소	대전시 유성구 어은동 99 한빛아파트 106동 1005호				
발명의 명칭		다층구조 신경망을 이용한 음소 분할 방법				

특허법 제 42 조의 규정에 의하여 위와 같이 출원합니다.

1995년 12 월 22 일

대리인 변리사 박 해 천

염 주 석

특 허 청 장 귀하

특허법 제 60 조의 규정에 의하여 위와 같이 출원심사를 청구합니다.

1995년 12 월 22 일

대리인 변리사 박 해 천

염 주 석

특 허 청 장 귀하

※ 구비서류

1. 출원서 부본 2통
2. 명세서, 요약서 및 도면 각3통
3. 위임장 1통

수수료			
출원료	기 본	20 면	20,000 원
	가 산	3 면	3,100 원
우선권 주장료		건	원
심사 청구료		1 항	75,000 원
합 계			98,100 원

명 세 서

1. 발명의 명칭

다층구조 신경망을 이용한 음소 분할 방법

2. 도면의 간단한 설명

제 1 도는 본 발명이 적용되는 시스템의 구성도,

제 2 도는 본 발명에 이용되는 다층 신경망의 구성도,

제 3 도는 본 발명의 일실시예에 따른 전체 흐름도,

* 도면의 주요부분에 대한 부호의 설명

1 : 음성 입력부

2 : 전처리부

3 : MLP 음소 분할부

4 : 음소 경계 출력부

3. 발명의 상세한 설명

본 발명은 다층구조 신경망을 이용한 음소 분할 방법에 관한 것이다.

종래의 음소 분할 기술은 음성 신호로부터 주파수 성분인 스펙트로그램(spectrogram)을 추출한 후 사전에 정해진 여러가지 음성학적 지식과 규칙을 적용한 분석을 통해 음소의 경계를 찾아내는 방법을 사용하는데, 이 방법은 우선 시스템의 복잡도가 매우 높다는 점이다. 또한 음소 분할을 위해 사용하는 여러가지 지식과 규칙 상호간의 효율적이고도 최적의 결합 방법이 없기 때문에 실제 사용시에는 시스템의 성능이 신뢰할 정도가 되지 못한다는 점과 실제 사용될 때의 상황 변화에 따라 급격히 저하된다는 문제점이 있다. 다른 방법으로는 음소 분할을 위해서 모든 음소들의 특징을 사전에 추출하여 패턴으로 저장한 후, 음소 분할시에 모든 음소들에 대한 특징 패턴을 번갈아가면서 입력된 음성 신호와 비교하여 음소의 경계를 찾아내는 기술을 들수 있다. 이 방법은 모든 음소들에 대한 특징 패턴의 정보를 가지고 있어야 하므로 시스템의 메모리 량이 커지게 되고 수행과정에서의 계산량도 증가함으로써 경제적인 시스템을 구현하지 못한다는 문제점이 있다.

따라서, 상기의 문제점을 해결하기 위하여 안출된 본 발명은 음소 자체에 대한 부가적인 지식없이 음소와 음소의 경계에서 나타나는 음성 신호상의 변화 만을 이용하여 음소의 경계가 되는 지점을 정확하고도 효율적으로 포착하여 자동 음소 분할이나 음소 레이블링이 필요한 응용 분야에 유익하게 활용될 수 있는 다층구조 신경망을 이용한 음소 분할 방법을 제공하는 데 그 목적이 있다.

상기의 목적을 달성하기 위한 발성된 음성으로 부터 디지탈로 변환된 음성 샘플을 출력하는 음성 입력부, 상기 음성 입력부로 부터 입력된 음성 샘플로 부터 음소 분할에 적합한 특징 벡터를 추출하는 전처리부, 상기 전처리부의 특징 벡터를 이용하여 음소의 경계 부분을 찾아 출력하는 다층 신경망(MLP : Multi Layer Perceptron) 음소 분할부, 및 상기 MLP 음소 분할부의 음소의 경계에 관한 위치 정보를 프레임 위치의 형태로 출력하는 음소 경계 출력부를 구비한 음소 분할 장치에 적용되는 다층구조 신경망을 이용한 음소 분할 방법에 있어서, 디지탈화된 음성 샘플들로부터 음성의 특징을 추출하기에 알맞은 길이로 음성을 연속적으로 세그먼트화하여 음성을 프레임화하고, 각 음성 프레임들에 대하여 음성학적인 지식에 근거하여 추출한 후 프레임별 특징 벡터들의 인접 프레임간 차이를 구한 프레임간 특징 벡터를 추출한 후 특징 벡터들의 최대

치와 최소치를 정규하여 다층 신경망(MLP)의 입력층과 은닉층, 은닉층과 출력층 사이에 존재하는 가중 함수를 초기화 한 후 다층 신경망(MLP)의 출력 목표 데이타를 지정하고, 특징 벡터를 다층 신경망(MLP)에 입력하여 오류 역전파 알고리즘을 사용하여 학습 시키 후, 평균자승 오차의 감소 비율이 허용 한계내로 수렴되면 학습을 통해 구한 가중 함수와 MLP의 규격에 대한 정보를 저장하고 종료하는 제 1 단계; 및 상기 제 1 단계 수행 후, 입력된 디지탈화된 음성 샘플들로부터 음성의 특징을 추출하기에 알맞은 길이로 음성을 연속적으로 세그먼트화하여 음성을 프레임화 한 후, 특징 벡터를 추출하고, 이를 정규화한 후, 상기 제 1 단계에서 구한 가중 함수를 읽고, 특징 벡터를 입력받아 음소 경계 판정을 위한 연산을 수행하여 출력값을 생성한 후, 출력값에 따라 음소 경계 부분을 판정하고, 현재의 분석 프레임이 입력된 음성의 최종 프레임의 2 프레임 이전에 도달하였으면, 구해진 음소의 경계를 프레임 번호로 나타낸 값을 최종 결과로 출력하는 제 2 단계를 포함하는 것을 특징으로 한다.

이하, 첨부된 도면을 참조하여 본 발명의 일실시예를 상세히 설명한다.

제 1 도는 본 발명이 적용되는 시스템의 구성도로서, 도면에서 1은

음성입력부, 2는 전처리부, 3은 다층 신경망(MLP : Multi Layer Perceptron) 음소 분할부, 4는 음소 경계 출력부를 각각 나타낸다.

음성입력부(1)는 공기중의 음성 파형으로 부터 전기적인 음성 신호로 변환하는 마이크와 전기적인 아날로그 신호로 입력된 음성 신호에서 저주파 잡음과 고주파 에일리어싱(aliasing) 효과를 제거하기 위한 대역통과 여파기(band pass filter), 및 아날로그 음성 신호를 디지탈 음성 신호로 변환하는 아날로그/디지탈 변환기(ADC : Analog to Digital Converter)로 구성되어 있으며 발성된 음성으로 부터 디지탈로 변환된 음성 샘플을 얻어 전처리부(2)에 출력한다.

상기 전처리부(2)는 상기 음성 입력부(1)에서 입력된 음성샘플들로 부터 음소 분할에 적합한 특징 벡터를 추출하여 MLP 음소 분할부(3)로 출력한다.

상기 MLP 음소 분할부(3)는 상기 전처리부(2)로 부터 입력된 특징 벡터들을 사용하여 음소의 경계 부분을 찾아 음소 경계 출력부(4)로 출력한다.

상기 음소 경계 출력부(4)는 상기 MLP 음소 분할부(3)에서 자동적으로 분할된 음소의 경계에 관한 위치 정보를 프레임 위치의 형태로 출력한다.

제 2 도는 본 발명에 이용되는 다층 신경망의 구성도를 나타낸다.

본 발명은 기존의 지식이나 규칙 기반의 음소분할 방법의 단점을 보완하기 위하여 효과적이고도 신뢰성있는 자동 음소 분할기(phoneme segmenter)를 신경망의 한 종류인 다층 신경망(MLP)을 사용하여 구현하였다. MLP를 이용하는 음소 분할 방법은 기존의 음소분할 방법들의 문제점으로 알려진 음성 신호에 내재되어 있는 음소 경계에 관한 지식이나 규칙의 불완전한 모델링에서 오는 성능 저하를 해결하는데 매우 적합하다. 즉, 많은 음성 데이타에서 추출한 특징 벡터로 부터 음소 분할에 필요한 기능을 학습을 통해서 스스로 배우도록 함으로써 음소의 경계에 대한 특별한 가정이나 규칙 및 지식을 사전에 도입하지 않고도 음성신호 자체에 내재된 지식이나 규칙을 MLP로 하여금 스스로 찾아내도록 하는 방법이다. 따라서, 이 방법에는 음성 신호의 모델링을 용이하게 하기 위하여 사전에 그 분포나 모델링을 위한 불확실한 가정의 도입이나 추가적인 처리를 할 필요가 없는 장점이 있다.

따라서, 본 발명에 이용되는 다층 신경망(MLP)의 구조는 입력(input), 은닉(hidden), 출력(output)의 세 가지 층(layer)으로 구성된 다층 구조의 형태를 취하고 있다. 도면에서와 같이 하단에 위치한 입력층은 연속적인 다섯 프레임에서 발생하는 4개의 인접 프레임 간 차이로 부터

추출된 총 72개의 프레임간 특징 벡터들에 대한 입력 노드들과 다층 신경망(MLP)의 은닉층에서의 문턱치 비교 과정 대신에 사용되는 입력값 1을 위한 입력 노드 한개를 포함하여 모두 73개의 입력 노드로 구성되어 있다. 출력층의 출력 노드는 음소의 경계임을 나타내는 첫번째 노드와 그렇지 않은 경우를 나타내는 두번째 노드를 합하여 모두 2개로 구성되어 있으며, 입력층과 출력층의 사이에 위치한 은닉층은 다층 신경망(MLP)이 실제로 구현해야 하는 비선형 분리(nonlinear discrimination)기능이 이루어지는 계층이다. 이 은닉층의 활성화 함수(activation function)로 다음과 같은 비선형의 S자 모양(sigmoid)의 함수를 사용한다.

$$y = (\exp(x) - 1) / (\exp(x) + 1),$$

여기서, x, y는 각각 활성화 함수의 입력과 출력을 나타낸다.

은닉층의 노드 수 N은 다층 신경망(MLP)의 최종 성능과 밀접한 관련이 있다고 알려져 있는데 여러가지 데이타를 사용한 실험을 통해서 10에서 30 사이가 적당하다.

입력층과 은닉층, 은닉층과 출력층 사이에는 각 층의 노드들을 전부

연결하는 가중함수(weight)들이 존재한다. 이 가중 함수들은 층과 층 사이의 노드들을 전부 연결시키기 때문에, 입력층과 은닉층의 경우에는 그 수가 입력노드의 수 x 은닉노드의 수 = 73 x N 개가 있으며, 은닉층과 출력층의 경우에는 은닉노드의 수 x 출력노드의 수 = N x 2 개가 존재한다. 이 가중 함수들은 오류 역전파 알고리즘을 이용한 학습을 통해서 사전에 구해진 다음 메모리에 저장되어 있다가 음소 분할시에 불러내어 사용된다.

제 3 도는 본 발명의 일실시예에 따른 전체 흐름도로서, 전처리부(2)와 MLP 음소 분할부(3)의 내부에서 음소 분할 알고리즘의 동작 과정을 나타낸 것으로 MLP 음소 분할 알고리즘의 학습 과정과 분할 과정의 2 부분으로 구성되어 있다.

먼저, 음성 프레임화와 특징 벡터 추출 과정은 전처리부(2)에서 수행되는 과정으로서 학습과 분할 두 과정에 공통적으로 사용된다. 본 발명에서의 특징 벡터들의 선정 에서는 음소간의 경계에서 음성 스펙트럼의 변화가 심하다는 점을 이용하기 위하여 각 프레임간의 스펙트럼의 차이를 잘 나타내주는 인자를 유도하였다.

먼저, 디지탈화된 음성 샘플들로부터 음성의 특징을 추출하기에 알맞은 길이로 음성을 연속적으로 세그먼트화하여 음성을 프레임화 한다

(10). 음성 프레임화는 입력된 전체 음성 샘플들에 대해서 매 10 msec마다 16 msec의 길이로 해밍(Hamming) 창함수(window)를 취하여 음성 프레임화한다.

다음은 음성 프레임으로 부터 특징 벡터 추출하는데 첫단계에서는 앞에서 구해진 각 음성 프레임들에 대하여 음성의 특징을 효과적으로 잘 나타내는 프레임별 특징 벡터들을 음성학적인 지식에 근거하여 추출하고, 두번째 단계에서는 첫단계에서 구한 프레임별 특징 벡터들의 인접 프레임간 차이를 구한 프레임간 특징 벡터를 추출하여 이들을 MLP 음소 분할부(3)에 입력되는 최종적인 특징 벡터로 사용한다(11).

먼저 1차적으로 각각의 프레임들에 대해서 구한 특징 벡터는 다음과 같다.

(1) 프레임 에너지 : 음성의 프레임별 발성강도를 나타내는 것으로서 다음과 같이 구하였다.

$$ENG\_FRM(t) = \log 10\left(\sum_{n} s(n)*s(n)\right) , \quad n=0,1, \ldots,N$$

,여기서 s(n)은 t번째 프레임에 속한 음성 샘플을 나타내고, N은 음성 프레임의 길이를 나타낸다.

(2) 16차 멜 스케일 FFT(mel-scaled fast Fourier transform) : 프레임별 음성의 주파수 특성인 스펙트럼을 구하기 위하여 먼저 FFT(fast Fourier transform)를 한 후 얻어진 음성의 주파수 성분을 인간의 청각 특성과 유사하게 사전에 정해진 16 개의 주파수 대역으로 분류한 16차의 대역별 에너지를 구하여 멜 스케일 FFT 계수로 사용한다. 프레임 인덱스 t에서 j차 멜 스케일 FFT 계수는 다음 식과 같이 구해진다.

$$MSFC(j, t) = \log10\left(\sum_{f=1}^{16} s(j,t,f)\right),$$ f는 각 주파수 밴드에 포함된 주파수

,여기서 j는 각 주파수 대역의 인덱스를 나타내고 s(j,t,f)는 FFT로부터 구해진 t번째 프레임의 j차 주파수 대역 진폭 스펙트럼의 주파수별 성분을 나타낸다.

(3) 대역별 에너지 비 : 음소 분할시에 유성음과 무성음으로 된 음소를 정확하게 구분하는 일이 매우 중요한데 이 유,무성음의 큰 차이점은 에너지의 주파수 대역별 분포이다. 따라서, 본 발명에서는 유,무성음의 구분을 위해 0-3kHz 사이에 존재하는 저주파 에너지와 3 kHz - 8 kHz 사이에 분포하는 고주파 에너지를 각각 구한다음 이들의 비를 특징 벡터의 하나로 선정하였다.

$$ENG\_RTO(t) = \log10(ENG\_LOW(t)) - \log10(ENG\_HIGH(t))$$

$$ENG\_LOW(t) = \sum_f s(f, t), \quad f=0,..., 3\ kHz.$$

$$ENG\_HIGH(t) = \sum_f s(f, t), \quad f=3kHz, ..., 8\ kHz.$$

,여기서 ENG_LOW(t), ENG_HIGH(t)는 각각 t 번째 음성 프레임의 저주파대와 고주파대의 에너지로서 FFT에서 구한 진폭 스펙트럼에서 각 대역에 포함된 성분들의 합으로 구한다.

최종적인 MLP 음소 분할부(3)의 입력으로 사용되는 프레임간 특징 벡터는 음소 분할이 음소간의 경계에서 큰 변화를 나타낸다는 특징에 근거하여 위에서 구한 일차적인 프레임별 특징 벡터들에 대해서 인접 프레임간의 차이를 다음과 같이 구함으로써 얻는다.

(1) 프레임 에너지의 인접 프레임간 차이

$$dENG\_FRM(t) = |ENG\_FRM(t) - ENG\_FRM(t-1)|$$

(2) 16차 멜 스케일 FFT의 프레임간 차이

$$dMSFC(j,t) = |MSFC(j,t) - MSFC(j, t-1)|, \quad j=0,1, ..., 15.$$

,여기서 j는 계수들의 각 차수를 나타낸다.

(3) 대역별 에너지 비의 프레임간 차이

$$dENG\_RTO(t) = |ENG\_RTO(t) - ENG\_RTO(t-1)|$$

이렇게 특징 벡터를 추출한 후, MLP 음소 분할부(3)의 입력으로 사용하기 위해 특징 벡터들의 최대치와 최소치가 각각 1과 -1이 되도록 정규화(normalize) 한다(12).

이렇게 정규화된 특징벡터를 이용한 MLP 음소 분할부(3)의 학습 과정을 살펴보면, MLP 음소 분할부(3)의 학습하기 위한 초기 단계로서 입력층과 은닉층, 은닉층과 출력층 사이에 존재하는 가중 함수(weight)를 초기화한다(13). 초기치는 1과 -1 사이에 분포하는 무작위의 값으로 선정하였다. 그런 다음 음소의 경계 부분을 찾도록 가르치는 출력층의 출력 목표 데이타를 지정한다(14). 각 프레임별로 출력 목표 데이타는 MLP 출력 노드의 수와 같은데, 음소의 경계일 경우(1,-1) 경계가 아닐

경우 (-1,1)의 값을 갖는다. 이 출력 목표 데이타는 사전에 음소 분할된 음성 데이타베이스로 부터 구한 음소의 경계 정보를 이용하여 해당 특징 벡터의 프레임 위치와 일치되도록 작성된다. 이렇게 출력 목표 데이타를 지정한 후, 학습 데이타인 특징 벡터를 MLP의 입력층에 입력하여(15), MLP를 학습 시킨다(16). 입력층에는 연속하는 4개의 프레임간 특징벡터의 입력을 위한 72개의 입력 노드와 은닉층의 문턱값 비교 과정 대신에 입력되는 1을 위한 하나의 입력 노드를 합하여 전체 73개의 노드로 구성된다. 4개의 프레임간 특징 벡터들은 제 2 도에 나타낸 하단에서와 같이 현재 분석 프레임 t를 중심으로 전후 2 프레임(t-2, t-1, t+1, t+2)씩을 포함한 5 프레임으로 부터 발생하는 4개의 프레임 사이에서 각각 추출된다. 음소분할 MLP의 학습과정이다. 학습 알고리즘은 일반적으로 사용하는 오류역전파(error back propagation)을 사용한다. 이렇게 MLP를 학습시킨 후, 평균자승오차(mean squared error)의 감소비율이 허용한계내로 수렴하였으면(17) 학습을 통해서 구해진 가중 함수들과 MLP의 규격에 대한 정보를 저장한 후(18) 학습 과정을 종료한다.

학습 과정을 종료한 후 상기에서 설명한 바와 같이 디지탈화된 음성 샘플들로부터 음성의 특징을 추출하기에 알맞은 길이로 음성을 연속적으로 세그먼트화하여 음성을 프레임화 한 후(10), 특징 벡터를 추출하

고(11), 이를 정규화 한다(12).

그런 다음 상기 학습 과정에서 구해진 가중 함수들을 MLP 은닉층으로 읽어들이고(19), 상기 과정에서 구한 특징벡터 72개를 MLP의 입력 노드에 순서에 따라서 입력시키고 마지막 73번째 입력노드에 1을 입력한다(20). MLP 음소 분할부(3)에서는 입력된 특징 벡터들에 대하여 다음과 같은 MLP 연산을 통해서 음소 경계 판정을 위한 출력값을 생성한다(21).

$$HID(j) = SGMOD(\sum_i IN(i) \times WGT\_IH(i,j)),\ i=0,1,...,72.\ j=0,1,...,N-2,$$

$$HID(N-1) = 1,$$

$$OUT(k) = SGMOD(\sum_j HID(j) \times WGT\_HO(j,k)),\ j=0,1,...,N-1,\ k=0,1.$$

여기서 IN(j)는 i 번째 입력 노드의 입력을, OUT(k)는 k번째 출력 노드의 출력을, WGT_IH(i,j)는 i번째 입력 노드와 j번째 은닉 노드를 연결하는 가중 함수를, WGT_HO(j,k)는 j번째 은닉 노드와 k번째 출력 노드를 연결하는 가중 함수를 나타내며, SGMOD는 전술한 S자 모양(sigmoid) 함수를 나타낸다. 또한 최종 출력 노드에서의 문턱값 비교 과

정을 대신하기 위해 마지막 은닉 노드에 1을 지정한다.

다음 음소 경계 부분을 판정하는 데 앞의 MLP 음소 분할부(3)에서 연산된 출력값을 비교하여 첫번째 출력값인 OUT(0)이 양수이면 그 분석 프레임이 음소의 경계이고 반대로 OUT(1)이 양수이면 음소의 경계가 아닌 것으로 판정한다(22).

그런 후, 현재의 분석 프레임이 입력된 음성의 최종 프레임의 2 프레임 이전에 도달하였나를 검사하여(23) 도달하지 않았으면 MLP 입력층에 특징 벡터를 입력하는 이하의 과정을 반복하고, 도달하였으면 구해진 음소의 경계를 프레임 번호로 나타낸 값을 최종 결과로 출력하고(24), 종료한다.

상기와 같이 동작하는 본 발명은 인간과 기계사이의 대화를 가능하게 해주는 기술인 음성 인식 시스템의 구현에 있어서 먼저 음소 단위로 음성을 분할한 다음 분할된 음소 세그먼트에 대하여 음소인식을 수행하는 음소분할 기반의 음소 인식에 필수적인 정확하고 효율적인 음소분할 전처리를 가능하게 해주며, 음소 단위의 음성 인식 및 음성합성 시스템의 구현에 필요한 다량의 음소 분할된 음성 데이타베이스를 구축할 때도 지금까지의 음성전문가에 의한 수작업을 대신하여 신뢰성과 일관성있게 자동적인 음성분할을 가능하게 함으로서 많은 시간과 비용의 절감을 가

져오는 효과가 있다.

4. 특허 청구의 범위

1. 발성된 음성으로 부터 디지탈로 변환된 음성 샘플을 출력하는 음성 입력부(1), 상기 음성 입력부(1)로 부터 입력된 음성 샘플로 부터 음소 분할에 적합한 특징 벡터를 추출하는 전처리부(2), 상기 전처리부(2)의 특징 벡터를 이용하여 음소의 경계 부분을 찾아 출력하는 다층 신경망(MLP : Multi Layer Perceptron) 음소 분할부(3), 및 상기 MLP 음소 분할부(3)의 음소의 경계에 관한 위치 정보를 프레임 위치의 형태로 출력하는 음소 경계 출력부(4)를 구비한 음소 분할 장치에 적용되는 다층 구조 신경망을 이용한 음소 분할 방법에 있어서,

디지탈화된 음성 샘플들로부터 음성의 특징을 추출하기에 알맞은 길이로 음성을 연속적으로 세그먼트화하여 음성을 프레임화하고, 각 음성 프레임들에 대하여 음성학적인 지식에 근거하여 추출한 후 프레임별 특징 벡터들의 인접 프레임간 차이를 구한 프레임간 특징 벡터를 추출한 후 특징 벡터들의 최대치와 최소치를 정규하여 다층 신경망(MLP)의 입력층과 은닉층, 은닉층과 출력층 사이에 존재하는 가중 함수를 초기화한 후 다층 신경망(MLP)의 출력 목표 데이타를 지정하고, 특징 벡터를 다층 신경망(MLP)에 입력하여 오류 역전파 알고리즘을 사용하여 학습

시키 후, 평균자승오차의 감소 비율이 허용 한계내로 수렴되면 학습을 통해 구한 가중 함수와 MLP의 규격에 대한 정보를 저장하고 종료하는 제 1 단계(10 내지 18); 및

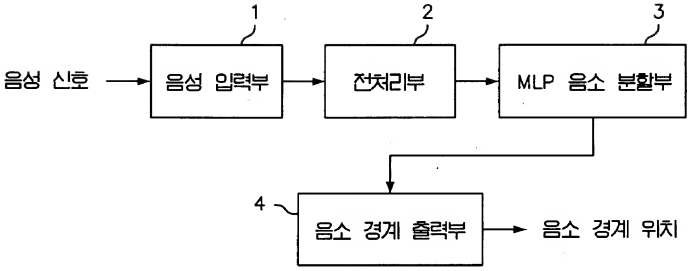
상기 제 1 단계(10 내지 18) 수행 후, 입력된 디지탈화된 음성 샘플들로부터 음성의 특징을 추출하기에 알맞은 길이로 음성을 연속적으로 세그먼트화하여 음성을 프레임화 한 후, 특징 벡터를 추출하고, 이를 정규화한 후, 상기 제 1 단계(10 내지 18)에서 구한 가중 함수를 읽고, 특징 벡터를 입력받아 음소 경계 판정을 위한 연산을 수행하여 출력값을 생성한 후, 출력값에 따라 음소 경계 부분을 판정하고, 현재의 분석 프레임이 입력된 음성의 최종 프레임의 2 프레임 이전에 도달하였으면, 구해진 음소의 경계를 프레임 번호로 나타낸 값을 최종 결과로 출력하는 제 2 단계(10,11,12,19 내지 24)를 포함하는 것을 특징으로 하는 다층구조 신경망을 이용한 음소 분할 방법.

요 약 서

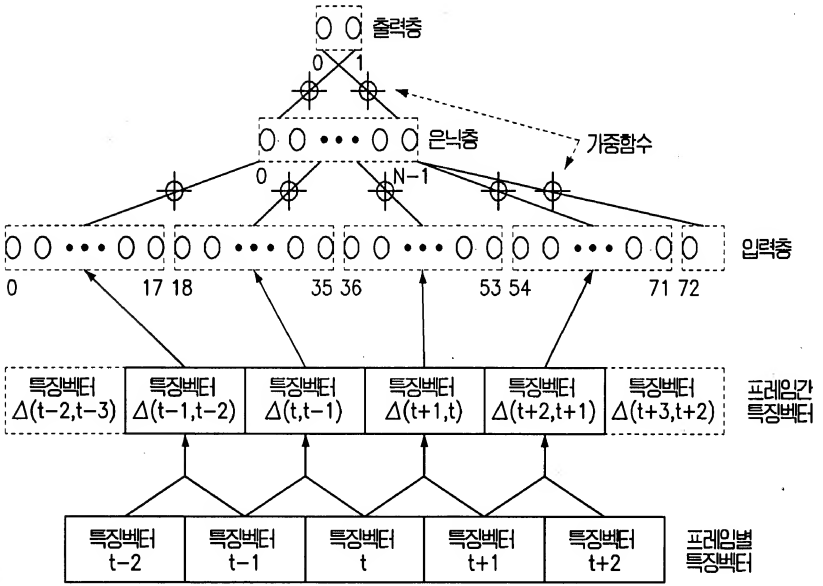
본 발명은 다층구조 신경망을 이용한 음소 분할 방법에 관한 것으로서, 음성 입력부(1), 전처리부(2) 다층 신경망(MLP : Multi Layer Perceptron) 음소 분할부(3), 및 음소 경계 출력부(4)를 구비한 음소 분할 장치에 적용되는 다층구조 신경망을 이용한 음소 분할 방법에 있어서, 디지탈화된 음성 샘플들로 부터 음성을 연속적으로 세그먼트화하여 음성을 프레임화하고, 각 음성 프레임들에 대하여 프레임간 특징 벡터를 추출한 후 특징 벡터들의 최대치와 최소치를 정규화하고, 가중 함수를 초기화 한 후 다층 신경망(MLP)의 출력 목표 데이타를 지정하고, 특징 벡터를 입력하여 오류 역전파 알고리즘을 사용하여 학습 시키 후, 평균자승오차의 감소 비율이 허용 한계내로 수렴되면 학습을 통해 구한 가중함수와 MLP의 규격에 대한 정보를 저장하고 종료하는 제 1 단계(10 내지 18); 및 상기 제 1 단계(10 내지 18) 수행 후, 음성을 프레임화 한 후, 특징 벡터를 추출하고, 이를 정규화한 후, 저장된 가중 함수를 읽고, 특징 벡터를 입력받아 음소 경계 판정을 위한 연산을 수행하여 출력값을 생성한 후, 출력값에 따라 음소 경계 부분을 판정하고, 현재의 분석 프레임이 입력된 음성의 최종 프레임의 2 프레임 이전에 도달하였으면, 구해

진 음소의 경계를 프레임 번호로 나타낸 값을 최종 결과로 출력하는 제 2 단계(10,11,12,19 내지 24)를 포함하여 음소 자체에 대한 부가적인 지식 없이 음소와 음소의 경계에서 나타나는 음성 신호상의 변화 만을 이용하여 음소의 경계가 되는 지점을 정확하고도 효율적으로 포착하여 자동 음소 분할이나 음소 레이블링이 필요한 응용 분야에 유익하게 활용될 수 있는 효과가 있다.

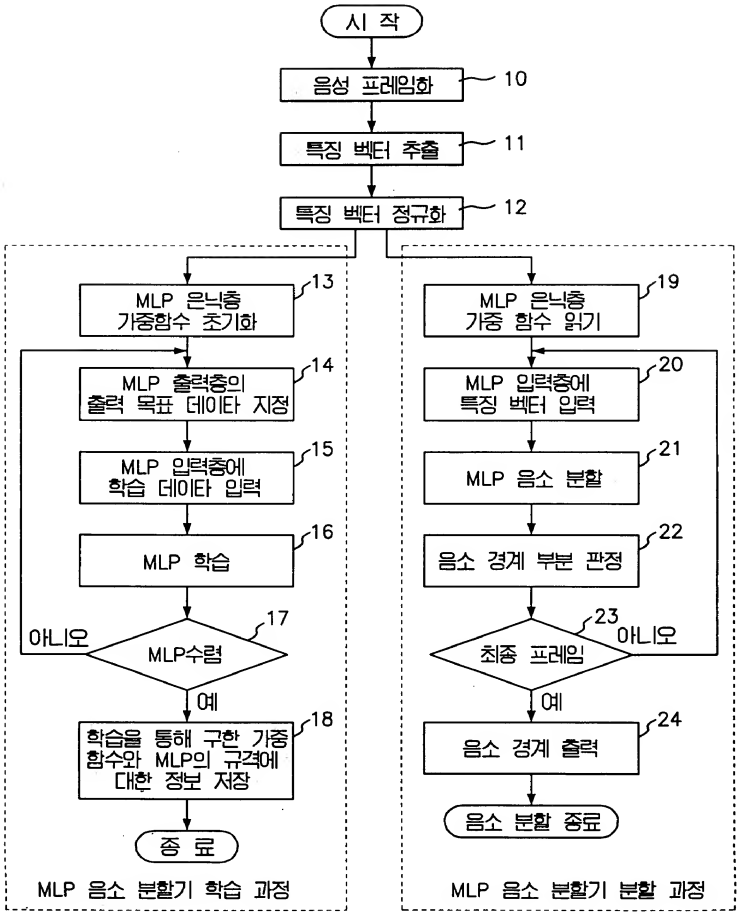
제 1 도



제 2 도


출력층
0
1
은닉층
가중함수
0
N-1
입력층
0
17
18
35
36
53
54
71
72
특징벡터 Δ(t-2,t-3)
특징벡터 Δ(t-1,t-2)
특징벡터 Δ(t,t-1)
특징벡터 Δ(t+1,t)
특징벡터 Δ(t+2,t+1)
특징벡터 Δ(t+3,t+2)
프레임간 특징벡터
특징벡터 t-2
특징벡터 t-1
특징벡터 t
특징벡터 t+1
특징벡터 t+2
프레임별 특징벡터

제 3 도


시 작
음성 프레임화
10
특징 벡터 추출
11
특징 벡터 정규화
12
MLP 은닉층 가중함수 초기화
13
MLP 출력층의 출력 목표 데이타 지정
14
MLP 입력층에 학습 데이타 입력
15
MLP 학습
16
아니오
MLP수렴
17
예
학습을 통해 구한 가중함수와 MLP의 규격에 대한 정보 저장
18
종 료
MLP 음소 분할기 학습 과정
MLP 은닉층 가중 함수 읽기
19
MLP 입력층에 특징 벡터 입력
20
MLP 음소 분할
21
음소 경계 부분 판정
22
최종 프레임
23
아니오
예
음소 경계 출력
24
음소 분할 종료
MLP 음소 분할기 분할 과정

위 임 장

변리사	성 명	박 해 천 염 주 석		변리사등록번호	F196 H386
	주 소	서울시 강남구 역삼1동 740-5 동방빌딩 1층			
사건의 표시		특허출원건			
발명의 명칭		다층구조 신경망을 이용한 음소 분할 방법			
위임자	성 명	한국전자통신연구소 소장 양 승 택		출원인코드	37500300
	주 소	대전직할시 유성구 가정동 161		우편번호	305-350
	사건과의 관계	출원인			
	성 명			출원인코드	
	주 소			우편번호	
	사건과의 관계				
위임할사항	1. 상기 사건에 관한 일체의 행위 및 이건에 관한 국내우선권 주장이나 그 취하, 출원심사의 청구, 우선심사신청, 간행물의 제출, 포기 혹은 취하, 출원인 명의변경, 출원분할, 출원변경, 증명의 청구, 이건의 사정, 심결 또는 처분에 대하여 항고심판, 즉시항고, 재심, 소원, 소송, 또는 상고를 처리하는 권한. 2. 상기 사항을 처리하기 위한 복대리인의 선임 및 해임에 관한 권한.				

특허법 제 7 조 실용신안법 제 3 조 의장법 제 4 조 및 상표법 제 5 조의 규정에 의한 절차를 위와 같이 위임함.

1995 년 12 월 18 일

위임자 한국전자통신연구소
소 장 양 승 택